

Fake News Detector

Installing Necessary Libraries

```
In [1]: import warnings
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import string
import re
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
warnings.filterwarnings('ignore')
```

Loading the data

```
In [2]: data_fake=pd.read_csv('Fake.csv')
data_true=pd.read_csv('True.csv')
```

Data Preview

```
In [3]: data_fake.head()
```

```
Out[3]:
```

	title	text	subject	date
0	Donald Trump Sends Out Embarrassing New Year's...	Donald Trump just couldn't wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017

```
In [4]: data_true.tail()
```

Out[4]:

	title	text	subject	date
21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017
21413	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews	August 22, 2017
21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017
21415	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017
21416	Indonesia to buy \$1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017

In [5]:

```
data_fake["class"] = 0
data_true['class'] = 1
```

In [6]:

```
data_fake.shape, data_true.shape
```

Out[6]:

```
((23481, 5), (21417, 5))
```

In [7]:

```
data_fake_manual_testing = data_fake.tail(10)
for i in range(23480, 23470, -1):
    data_fake.drop([i], axis = 0, inplace = True)

data_true_manual_testing = data_true.tail(10)
for i in range(21416, 21406, -1):
    data_true.drop([i], axis = 0, inplace = True)
```

In [8]:

```
data_fake.shape, data_true.shape
```

Out[8]:

```
((23471, 5), (21407, 5))
```

In [9]:

```
data_fake_manual_testing['class'] = 0
data_true_manual_testing['class'] = 1
```

In [10]:

```
data_fake_manual_testing.head(10)
```

Out[10]:

		title	text	subject	date	class
23471	Seven Iranians freed in the prisoner swap have...	21st Century Wire says This week, the historic...	Middle-east	January 20, 2016	0	
23472	#Hashtag Hell & The Fake Left	By Dady Chery and Gilbert MercierAll writers ...	Middle-east	January 19, 2016	0	
23473	Astro turfing: Journalist Reveals Brainwashing ...	Vic Bishop Waking TimesOur reality is carefull...	Middle-east	January 19, 2016	0	
23474	The New American Century: An Era of Fraud	Paul Craig RobertsIn the last years of the 20t...	Middle-east	January 19, 2016	0	
23475	Hillary Clinton: 'Israel First' (and no peace ...	Robert Fantina CounterpunchAlthough the United...	Middle-east	January 18, 2016	0	
23476	McPain: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	January 16, 2016	0	
23477	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	21st Century Wire says It s a familiar theme. ...	Middle-east	January 16, 2016	0	
23478	Sunnistan: US and Allied 'Safe Zone' Plan to T...	Patrick Henningsen 21st Century WireRemember ...	Middle-east	January 15, 2016	0	
23479	How to Blow \$700 Million: Al Jazeera America F...	21st Century Wire says Al Jazeera America will...	Middle-east	January 14, 2016	0	
23480	10 U.S. Navy Sailors Held by Iranian Military ...	21st Century Wire says As 21WIRE predicted in ...	Middle-east	January 12, 2016	0	

In [11]: `data_true_manual_testing.head(10)`

Out[11]:

	title	text	subject	date	class
21407	Mata Pires, owner of embattled Brazil builder ...	SAO PAULO (Reuters) - Cesar Mata Pires, the ow...	worldnews	August 22, 2017	1
21408	U.S., North Korea clash at U.N. forum over nuc...	GENEVA (Reuters) - North Korea and the United ...	worldnews	August 22, 2017	1
21409	U.S., North Korea clash at U.N. arms forum on ...	GENEVA (Reuters) - North Korea and the United ...	worldnews	August 22, 2017	1
21410	Headless torso could belong to submarine journ...	COPENHAGEN (Reuters) - Danish police said on T...	worldnews	August 22, 2017	1
21411	North Korea shipments to Syria chemical arms a...	UNITED NATIONS (Reuters) - Two North Korean sh...	worldnews	August 21, 2017	1
21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017	1
21413	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews	August 22, 2017	1
21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017	1
21415	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017	1
21416	Indonesia to buy \$1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017	1

In [12]: `data_merge=pd.concat([data_fake, data_true], axis = 0)`
`data_merge.head(10)`

Out[12]:

	title	text	subject	date	class
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn't wish all Americans ...	News	December 31, 2017	0
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	0
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	0
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	0
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	0
5	Racist Alabama Cops Brutalize Black Boy While...	The number of cases of cops brutalizing and ki...	News	December 25, 2017	0
6	Fresh Off The Golf Course, Trump Lashes Out A...	Donald Trump spent a good portion of his day a...	News	December 23, 2017	0
7	Trump Said Some INSANELY Racist Stuff Inside ...	In the wake of yet another court decision that...	News	December 23, 2017	0
8	Former CIA Director Slams Trump Over UN Bully...	Many people have raised the alarm regarding th...	News	December 22, 2017	0
9	WATCH: Brand-New Pro-Trump Ad Features So Muc...	Just when you might have thought we'd get a br...	News	December 21, 2017	0

"title", "subject" and "date" columns is not required for detecting the fake news, so I am going to drop the columns.

In [13]: `data_merge.columns`Out[13]: `Index(['title', 'text', 'subject', 'date', 'class'], dtype='object')`In [14]: `data = data_merge.drop(['title', 'subject', 'date'], axis=1)`In [15]: `#count of missing values
data.isnull().sum()`Out[15]: `text 0
class 0
dtype: int64`

Randomly shuffling the dataframe

In [16]: `data = data.sample(frac=1)`In [17]: `data.head()`

	text	class
15529	Well, it's not like we didn't know this but it...	0
13241	Like we didn't know this already BUT now we ha...	0
15834	Of course Emily's List is going to support Hil...	0
9140	WASHINGTON (Reuters) - U.S. Senate Majority Le...	1
21017	XIAMEN, China (Reuters) - BRICS countries shou...	1

```
In [18]: data.reset_index(inplace = True)
data.drop(['index'], axis = 1, inplace = True)
```

```
In [19]: data.columns
```

```
Out[19]: Index(['text', 'class'], dtype='object')
```

```
In [20]: data.head()
```

```
Out[20]:
```

	text	class
0	Well, it s not like we didn t know this but it...	0
1	Like we didn t know this already BUT now we ha...	0
2	Of course Emily s List is going to support Hil...	0
3	WASHINGTON (Reuters) - U.S. Senate Majority Le...	1
4	XIAMEN, China (Reuters) - BRICS countries shou...	1

Preprocessing Text

Creating a function to convert the text in lowercase, remove the extra space, special chr., url and links.

```
In [21]: def wordopt(text):
    text = text.lower()
    text = re.sub('.*?\]', ' ', text)
    text = re.sub("\W", " ", text)
    text = re.sub('https?://\S+|www\.\S+', '', text)
    text = re.sub('<.*?>+', b'', text)
    text = re.sub('[%s]' % re.escape(string.punctuation), ' ', text)
    text = re.sub('\w*\d\w*', ' ', text)
    return text
```

```
In [22]: data['text'] = data['text'].apply(wordopt)
```

Defining dependent and independent variable as x and y

```
In [23]: x = data['text']
y = data['class']
```

Training the model

Splitting the dataset into training set and testing set.

```
In [24]: x_train, x_test, y_train, y_test = train_test_split(x,y,test_size = 0.25)
```

Extracting Features from the Text

Convert text to vectors

```
In [25]: from sklearn.feature_extraction.text import TfidfVectorizer  
  
vectorization = TfidfVectorizer()  
xv_train = vectorization.fit_transform(x_train)  
xv_test = vectorization.transform(x_test)
```

Logistic Regression

```
In [26]: from sklearn.linear_model import LogisticRegression
```

```
In [27]: LR = LogisticRegression()  
LR.fit(xv_train, y_train)
```

```
Out[27]: LogisticRegression(?)  
LogisticRegression()
```

```
In [28]: pred_lr = LR.predict(xv_test)
```

```
In [29]: LR.score(xv_test, y_test)
```

```
Out[29]: 0.9872549019607844
```

```
In [30]: print(classification_report(y_test, pred_lr))
```

	precision	recall	f1-score	support
0	0.99	0.99	0.99	5903
1	0.99	0.99	0.99	5317
accuracy			0.99	11220
macro avg	0.99	0.99	0.99	11220
weighted avg	0.99	0.99	0.99	11220

Decision Tree Classifier

```
In [31]: from sklearn.tree import DecisionTreeClassifier  
  
DT = DecisionTreeClassifier()  
DT.fit(xv_train, y_train)
```

Out[31]:

```
▼ DecisionTreeClassifier ⓘ ?  
DecisionTreeClassifier()
```

In [32]:

```
pred_dt = DT.predict(xv_test)
```

In [33]:

```
DT.score(xv_test, y_test)
```

Out[33]:

```
0.9967023172905526
```

In [34]:

```
print(classification_report(y_test, pred_lr))
```

	precision	recall	f1-score	support
0	0.99	0.99	0.99	5903
1	0.99	0.99	0.99	5317
accuracy			0.99	11220
macro avg	0.99	0.99	0.99	11220
weighted avg	0.99	0.99	0.99	11220

Gradient Boost Classifier

In [35]:

```
from sklearn.ensemble import GradientBoostingClassifier  
  
GB = GradientBoostingClassifier(random_state = 0)  
GB.fit(xv_train, y_train)
```

Out[35]:

```
▼ GradientBoostingClassifier ⓘ ?  
GradientBoostingClassifier(random_state=0)
```

In [36]:

```
pred_gb = GB.predict(xv_test)
```

In [39]:

```
GB.score(xv_test, y_test)
```

Out[39]:

```
0.9964349376114082
```

In [40]:

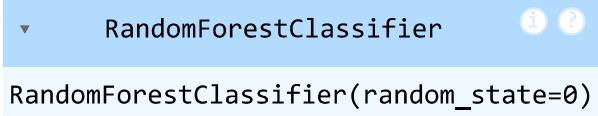
```
print(classification_report(y_test, pred_gb))
```

	precision	recall	f1-score	support
0	1.00	0.99	1.00	5903
1	0.99	1.00	1.00	5317
accuracy			1.00	11220
macro avg	1.00	1.00	1.00	11220
weighted avg	1.00	1.00	1.00	11220

Random Forest Classifier

```
In [41]: from sklearn.ensemble import RandomForestClassifier
RF = RandomForestClassifier(random_state = 0)
RF.fit(xv_train, y_train)
```

Out[41]:



```
In [42]: pred_rf = RF.predict(xv_test)
```

```
In [43]: RF.score(xv_test, y_test)
```

Out[43]: 0.9880570409982175

```
In [44]: print (classification_report(y_test, pred_rf))
```

	precision	recall	f1-score	support
0	0.99	0.99	0.99	5903
1	0.99	0.99	0.99	5317
accuracy			0.99	11220
macro avg	0.99	0.99	0.99	11220
weighted avg	0.99	0.99	0.99	11220

Testing the Model

```
In [45]: def output_label(n):
    if n==0:
        return "Fake News"
    elif n==1:
        return "Not A Fake News"

def manual_testing(news):
    testing_news = {"text":[news]}
    new_def_test = pd.DataFrame(testing_news)
    new_def_test['text'] = new_def_test['text'].apply(wordopt)
    new_x_test = new_def_test['text']
    new_xv_test = vectorization.transform(new_x_test)
    pred_LR = LR.predict(new_xv_test)
    pred_DT = DT.predict(new_xv_test)
    pred_GB = GB.predict(new_xv_test)
    pred_RF = RF.predict(new_xv_test)

    return print("\n\nLR Prediction: {} \nDT Prediction: {} \nGBC Prediction: {} \nRFC Prediction:{}".format(output_label(pred_LR[0]),
                                                                                                         output_label(pred_DT[0]),
```

```
output_label(pred_GB[0]),  
output_label(pred_RF[0]))
```

Model Testing With Manual Entry

```
In [46]: news = str(input())  
manual_testing(news)
```

LR Prediction: Not A Fake News
DT Prediction: Not A Fake News
GBC Prediction: Not A Fake News
RFC Prediction:Not A Fake News

```
In [47]: news=str(input())  
manual_testing(news)
```

LR Prediction: Fake News
DT Prediction: Fake News
GBC Prediction: Fake News
RFC Prediction:Fake News

```
In [49]: news=str(input())  
manual_testing(news)
```

LR Prediction: Not A Fake News
DT Prediction: Not A Fake News
GBC Prediction: Not A Fake News
RFC Prediction:Not A Fake News

```
In [51]: news=str(input())  
manual_testing(news)
```

LR Prediction: Fake News
DT Prediction: Fake News
GBC Prediction: Fake News
RFC Prediction:Fake News

```
In [ ]:
```